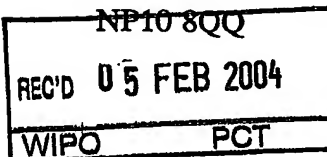**The Patent Office**

PCT/GB 2003 / 0 0 5 3 3 4

INVESTOR IN PEOPLE

# PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

REC'D 0 5 FEB 2004

WIPO        PCT

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.
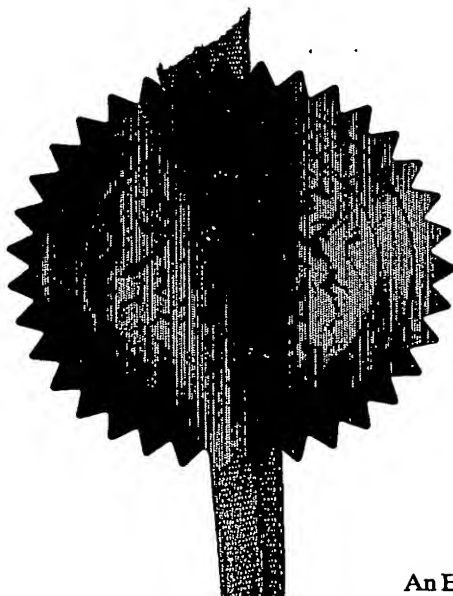
In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed *Stephen Hordley*

Dated    5 January 2004

Patents Form 1/77

Patents Act 1977
(Rule 16)

# Request for grant of a patent

*(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)*

| | | |
|---|---|---|
| 1. | Your reference | SW/P101151GB |

| | | | |
|---|---|---|---|
| 2. | Patent application number *(The Patent Office will fill in this part)* | **0228337.2** | **05 DEC 2002** |

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

Yorkshire Cancer Research
39 East Parade
Harrogate
HG1 5LQ

Patents ADP number *(if you know it)*    599857 0001

If the applicant is a corporate body, give the country/state of its incorporation

4. Title of the invention    REPLICATION PROTEIN

5. Name of your agent *(if you have one)*    Harrison Goddard Foote

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

Belgrave Hall
Belgrave Street
Leeds
LS2 8DD

Patents ADP number *(if you know it)*    ~~1497 1001~~ 763131 0002

| | | Country | Priority application number *(if you know it)* | Date of filing *(day / month / year)* |
|---|---|---|---|---|
| 6. | If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number | | | |

| | | Number of earlier application | Date of filing *(day / month / year)* |
|---|---|---|---|
| 7. | If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application | | |

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) *any applicant named in part 3 is not an inventor, or*    Yes

b) *there is an inventor who is not named as an applicant, or*

c) *any named applicant is a corporate body.*

*See note (d))*

Patents Form 1/77

**Patents Form 1/77**

9. Enter the number of sheets for any of the
   following items you are filing with this form.
   Do not count copies of the same document

| | |
|---|---|
| Continuation sheets of this form | |
| Description | 27 |
| Claim(s) | |
| Abstract | |
| Drawing(s) | 12 |

10. If you are also filing any of the following,
    state how many against each item.

Priority documents

Translations of priority documents

Statement of inventorship and right
to grant of a patent *(Patents Form 7/77)*

Request for preliminary examination
and search *(Patents Form 9/77)*

Request for substantive examination
*(Patents Form 10/77)*

Any other documents
*(please specify)*

11.                                        I/We request the grant of a patent on the basis of this application.

| Signature | Date |
|---|---|
| Harrison Goddard Foote | 4 December 2002 |

12. Name and daytime telephone number of
    person to contact in the United Kingdom

Siobhan Ward                                0113 233 0100

**Warning**

*After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication
or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You
will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the
United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting
written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the
United Kingdom for a patent for the same invention and either no direction prohibiting publication or
communication has been given, or any such direction has been revoked.*

**Notes**

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate
   sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be
   attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

**Patents Form 1/77**

# REPLICATION PROTEIN

This invention relates to a screening method for the identification of agents which modulate the activity of a DNA replication protein as a target for intervention in cancer therapy and includes agents which modulate said activity.

The DNA replication initiation process involves assembly of replication proteins into higher order complexes inside the nucleus during G1 phase of the cell cycle, followed by their activation to begin DNA synthesis (S phase). Intensive study has focused on a few proteins, largely identified by yeast genetics, that are involved in the replication complex assembly process and its regulation (Blow, 2001; Diffley and Labib, 2002). These include the origin recognition complex (ORC), Cdc6 and Cdt1, which are components of the pre-replication complex, and the Mcm proteins, which are believed to be the main replicative DNA helicase. The production and assembly of these proteins is regulated by cyclin dependent protein kinase 2 (cdk2) and by proteins which impinge on cdk2 activity, such as the cdk inhibitors p27$^{kip1}$ and p21$^{cip1}$.

Initiation of DNA replication can now be reconstituted with isolated mammalian nuclei and cytosolic extracts (Krude, 2000; Krude et al., 1997; Laman et al., 2001; Stoeber et al., 1998). Replication complex assembly and activation of DNA synthesis have been separated and reproduced under the regulation of recombinant cyclin-dependent kinases (cdks), using nuclei and extracts derived from G1 phase mouse cells (Coverley et al., 2002).

The mammalian initiation process can be separated into an assembly phase which is positively regulated by cyclin E -cdk2 and negatively regulated by cyclin A-cdk2, and an activation stage which is regulated by cyclin A -cdk2 (Coverley et al., 2002). Both phases can be reconstituted *in vitro* and are highly sensitive to recombinant cdk2 concentration. At the active concentration cyclin E-cdk2 stimulates replication complex assembly by cooperating with Cdc6, making G1 nuclei competent to replicate *in vitro*. In

contrast, cyclin A-cdk2 has two separable functions with *in vitro* optima at different concentrations: activation of DNA synthesis in replication complexes that are already assembled, and inhibition of assembly of new complexes. The dual functions of cyclin A ensure that the assembly phase (G1) ends before DNA synthesis (S) begins, thereby

5    preventing re-initiation until the next cell cycle.

A number of changes to chromatin bound proteins occur when DNA synthesis is activated *in vitro* by recombinant cyclin A-cdk2. The present invention relates to the finding that a cdc6-related antigen, p85, correlates with the initiation of DNA replication

10   and is regulated by cyclin A-cdk2. The protein was cloned from a mouse embryo library and identified as mouse Ciz1.

Human Ciz1 (Cip1 Interacting Zinc-finger protein) was described in 1999 after a two-hybrid screen to identify cyclin E/p21 complex interacting proteins. Mitsui et al (Mitsui

15   et al., 1999) showed that Ciz1 interacts with the cdk inhibitor p21$^{Cip1}$, but not with cyclin E. No analysis of Ciz1 function was reported, except that a role in transcription was sought but not found.

In vitro analysis has shown that Ciz1 protein positively regulates initiation of DNA

20   replication and that its activity is modulated by cdk phosphorylation at threonine 191/2, linking it to the cdk-dependent pathways that control initiation. The transcription factor-like features of Ciz1 are not required for replication function, implying that Ciz1 has more than one function. The Embryonic form mouse Ciz1 is alternately spliced, compared to predicted and somatic forms. Human Ciz1 is also alternately spliced, with

25   variability in the same exons as mouse Ciz1. It has been found that recombinant embryonic form Ciz1 promotes initiation of mammalian DNA replication and that pediatric cancers express 'embryonic-like' forms of Ciz1. Without wishing to be held to one theory, the inventors propose that Ciz1 mis-splicing produces embryonic form Ciz1 at inappropriate times in development. This promotes DNA replication and contributes

30   to formation or progression of cancer cell lineages.

2

A number of techniques have been developed in recent years which purport to specifically ablate genes and/or gene products. For example, the use of anti-sense nucleic acid molecules to bind to and thereby block or inactivate target mRNA molecules is an effective means to inhibit the production of gene products.

5

A much more recent technique to specifically ablate gene function is through the introduction of double stranded RNA, also referred to as inhibitory RNA (RNAi), into a cell which results in the destruction of mRNA complementary to the sequence included in the RNAi molecule. The RNAi molecule comprises two complementary strands of

10 RNA (a sense strand and an antisense strand) annealed to each other to form a double stranded RNA molecule. The RNAi molecule is typically derived from the exonic or coding sequence of the gene which is to be ablated.

Nucleic acids and proteins have both a linear sequence structure, as defined by their base

15 or amino acid sequence, and also a three dimensional structure which in part is determined by the linear sequence and also the environment in which these molecules are located. Conventional therapeutic molecules are small molecules, for example, peptides, polypeptides, or antibodies, which bind target molecules to produce an agonistic or antagonistic effect. It has become apparent that nucleic acid molecules also

20 have potential with respect to providing agents with the requisite binding properties which may have therapeutic utility. These nucleic acid molecules are typically referred to as aptamers. Aptamers are small, usually stabilised, nucleic acid molecules which comprise a binding domain for a target molecule.

25 Aptamers may comprise at least one modified nucleotide base. The term "modified nucleotide base" encompasses nucleotides with a covalently modified base and/or sugar. For example, modified nucleotides include nucleotides having sugars which are covalently attached to low molecular weight organic groups other than a hydroxyl group at the 3' position and other than a phosphate group at the 5' position. Thus modified

30 nucleotides may also include 2' substituted sugars such as 2'-O-methyl-; 2-O-alkyl; 2-O-

allyl; 2'-S-alkyl; 2'-S-allyl; 2'- fluoro-; 2'-halo or 2;azido-ribose, carbocyclic sugar analogues a-anomeric sugars; epimeric sugars such as arabinose, xyloses or lyxoses, pyranose sugars, furanose sugars, and sedoheptulose.

5    Modified nucleotides are known in the art and include by example and not by way of limitation; alkylated purines and/or pyrimidines; acylated purines and/or pyrimidines; or other heterocycles. These classes of pyrimidines and purines are known in the art and include, pseudoisocytosine; N4, N4-ethanocytosine; 8-hydroxy-N6-methyladenine; 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil; 5-fluorouracil; 5-bromouracil; 5-

10   carboxymethylaminomethyl-2-thiouracil;    5-carboxymethylaminomethyl   uracil; dihydrouracil; inosine; N6-isopentyl-adenine; 1-methyladenine; 1-methylpseudouracil; 1-methylguanine;    2,2-dimethylguanine;    2-methyladenine;    2-methylguanine;    3-methylcytosine;    5-methylcytosine;    N6-methyladenine;    7-methylguanine;    5-methylaminomethyl    uracil;    5-methoxy    amino    methyl-2-thiouracil;    β-D-

15   mannosylqueosine; 5-methoxycarbonylmethyluracil; 5-methoxyuracil; 2 methylthio-N6-isopentenyladenine; uracil-5-oxyacetic acid methyl ester; psueouracil; 2-thiocytosine; 5-methyl-2 thiouracil, 2-thiouracil; 4-thiouracil; 5-methyluracil; N-uracil-5-oxyacetic acid methylester; uracil 5—oxyacetic acid; queosine; 2-thiocytosine; 5-propyluracil; 5-propylcytosine;    5-ethyluracil;    5-ethylcytosine;    5-butyluracil;    5-pentyluracil;    5-

20   pentylcytosine; and 2,6,-diaminopurine; methylpsuedouracil; 1-methylguanine; 1-methylcytosine;

Aptamers may be synthesized using conventional phosphodiester linked nucleotides using standard solid or solution phase synthesis techniques which are known in the art.

25   Linkages between nucleotides may use alternative linking molecules. For example, linking groups of the formula P(O)S, (thioate); P(S)S, (dithioate); P(O)NR'2; P(O)R'; P(O)OR6; CO; or CONR'2 wherein R is H (or a salt) or alkyl (1-12C) and R6 is alkyl (1-9C) is joined to adjacent nucleotides through –O- or –S-.

4

Other techniques which purport to specifically ablate genes and/or gene products focus on modulating the function or interfering with the activity of protein molecules. Proteins can be targeted by chemical inhibitors drawn, for example, from existing small molecule libraries.

5

Antibodies, preferably monoclonal, can be raised for example in mice or rats against different protein isoforms. Antibodies, also known as immunoglobulins, are protein molecules which have specificity for foreign molecules (antigens). Immunoglobulins (Ig) are a class of structurally related proteins consisting of two pairs of polypeptide

10 chains, one pair of light (L) (low molecular weight) chain ($\kappa$ or $\lambda$), and one pair of heavy (H) chains ($\gamma$, $\alpha$, $\mu$, $\delta$ and $\epsilon$), all four linked together by disulphide bonds. Both H and L chains have regions that contribute to the binding of antigen and that are highly variable from one Ig molecule to another. In addition, H and L chains contain regions that are non-variable or constant.

15

The L chains consist of two domains. The carboxy-terminal domain is essentially identical among L chains of a given type and is referred to as the "constant" (C) region. The amino terminal domain varies from one L chain to anther and contributes to the binding site of the antibody. Because of its variability, it is referred to as the "variable" (V) region.

20 (V) region.

The H chains of Ig molecules are of several classes, $\alpha$, $\mu$, $\sigma$, $\alpha$, and $\gamma$ (of which there are several sub-classes). An assembled Ig molecule consisting of one or more units of two identical H and L chains, derives its name from the H chain that it possesses. Thus,

25 there are five Ig isotypes: IgA, IgM, IgD, IgE and IgG (with four sub-classes based on the differences in the H chains, i.e., IgG1, IgG2, IgG3 and IgG4). Further detail regarding antibody structure and their various functions can be found in, Using Antibodies: A laboratory manual, Cold Spring Harbour Laboratory Press.

Chimeric antibodies are recombinant antibodies in which all of the V-regions of a mouse or rat antibody are combined with human antibody C-regions. Humanised antibodies are recombinant hybrid antibodies which fuse the complimentarity determining regions from a rodent antibody V-region with the framework regions from the human antibody V-

5    regions. The C-regions from the human antibody are also used. The complimentarity determining regions (CDRs) are the regions within the N-terminal domain of both the heavy and light chain of the antibody to where the majority of the variation of the V-region is restricted. These regions form loops at the surface of the antibody molecule. These loops provide the binding surface between the antibody and antigen.

10

Antibodies from non-human animals provoke an immune response to the foreign antibody and its removal from the circulation. Both chimeric and humanised antibodies have reduced antigenicity when injected to a human subject because there is a reduced amount of rodent (i.e. foreign) antibody within the recombinant hybrid antibody, while

15   the human antibody regions do not illicit an immune response. This results in a weaker immune response and a decrease in the clearance of the antibody. This is clearly desirable when using therapeutic antibodies in the treatment of human diseases. Humanised antibodies are designed to have less "foreign" antibody regions and are therefore thought to be less immunogenic than chimeric antibodies.

20

Other techniques for targetting at the protein level include the use of randomly generated peptides that specifically bind to proteins, and any other molecules which bind to proteins or protein variants and modify the function thereof.

25   Understanding the DNA replication process is of prime concern in the field of cancer therapy. It is known that cancer cells can become resistant to chemotherapeutic agents and can evade detection by the immune system. There is an on going need to identify targets for cancer therapy so that new agents can be identified. The DNA replication process represents a prime target for drug intervention in cancer therapy. There is a

30   need to identify gene products which modulate DNA replication and which contribute to

formation or progression of cancer cell lineages, and to develop agents that affect their function.

According to one aspect of the present invention there is provided the use of a
5    polypeptide with the activity of Ciz 1, or any variant thereof as a target for the identification of agents which modulate DNA replication.

According to an alternative aspect of the invention there is provided a screening method for the identification of agents which modulate DNA replication wherein the screening
10    method comprises the use of Ciz1 and variants thereof.

Preferably the screening method comprises the steps of:
(i)    forming a preparation comprising a polypeptide molecule, or an active fragment thereof, encoded by a nucleic acid molecule selected from the group consisting of:
15    a) a nucleic acid molecule comprising a nucleic acid sequence represented in Fig 8 a or b;
b) a nucleic acid molecule which hybridizes to the nucleic acid sequence in (a) and which has Ciz1 activity or activity of a variant thereof;
c)a nucleic acid molecule which has a nucleic acid sequence which is degenerate
20    because of the genetic code to the sequences in a) and b) and a candidate agent to be tested;
d) a nucleic acid molecule derived from the genomic sequence at the Ciz1 locus and
ii)    detecting or measuring the effect of the agent on the activity of said polypeptide.
25

Assays for the detection of DNA replication are known in the art. Activity residing in Ciz1, or derived peptide fragments, and the effect of potential therapeutic agents on that activity would be assayed *in vitro* or *in vivo*.

*In vitro* assays for Ciz1 protein activity would comprise synchronised isolated G1 phase nuclei and either S phase extract or G1 phase extract supplemented with cyclin-dependent kinases. Inclusion of Ciz1 or derived peptide fragments stimulates initiation of DNA replication in these circumstances and can be monitored visually (by scoring

5    nuclei that have incorporated fluorescent nucleotides during *in vitro* reactions) or by measuring incorporation of radioactive nucleotides. The assay for therapeutic reagents that interfere with Ciz1 protein function would involve looking for inhibition of DNA replication in these assays. The effect of agents on Ciz1 nuclear localisation, chromatin binding, stability, modification and protein-protein interactions could also be monitored

10   in these assays.


*In vivo* assays will include creation of cell and mouse models that over-express or under-express Ciz1, or derived fragments, resulting in altered cell proliferation. The preparation of transgenic animals is generally known in the art and within the ambit of

15   the skilled person. The assay for therapeutic reagents would involve analysis of cell-cycle time, initiation of DNA replication and cancer incidence in the presence and absence of drugs that either impinge on Ciz1 protein activity, or interfere with Ciz1 production by targeting Ciz1 and its variants at the RNA level.


20   In a preferred method of the invention said hybridisation conditions are stringent.


Stringent hybridisation/washing conditions are well known in the art. For example, nucleic acid hybrids that are stable after washing in 0.1xSSC,0.1% SDS at 60°C. It is well known in the art that optimal hybridisation conditions can be calculated if the

25   sequence of the nucleic acid is known. Typically, hybridisation conditions uses 4 – 6 x SSPE (20x SSPE contains 175.3g NaCl, 88.2g $NaH_2PO_4 H_2O$ and 7.4g EDTA dissolved to 1 litre and the pH adjusted to 7.4); 5-10x Denhardts solution (50x Denhardts solution contains 5g Ficoll (Type 400, Pharmacia), 5g polyvinylpyrrolidone abd 5g bovine serum albumen; 100µg-1.0mg/ml sonicated salmon/herring DNA; 0.1-1.0% sodium dodecyl

30   sulphate; optionally 40-60% deionised formamide. Hybridisation temperature will vary

depending on the GC content of the nucleic acid target sequence but will typically be between $42^0$- $65^0$ C.

In a preferred method of the invention said polypeptide is modified by deletion, substitution or addition of at least one amino acid residue of the sequence.

A modified or variant, i.e. a fragment polypeptide and reference polypeptide, may differ in amino acid sequence by one or more substitutions, additions, deletions, truncations which may be present in any combination. Among preferred variants are those that vary from a reference polypeptide by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid by another amino acid of like characteristics. The following non-limiting list of amino acids are considered conservative replacements (similar): a) alanine, serine, and threonine; b) glutamic acid and asparatic acid; c) asparagine and glutamine d) arginine and lysine; e) isoleucine, leucine, methionine and valine and f) phenylalanine, tyrosine and tryptophan. Preferred are variants which retain the same biological function and activity as the reference polypeptide from which it varies. Alternatively, variants include those with an altered biological function, for example variants which act as antagonists, so called "dominant negative" variants.

Alternatively or in addition, non-conservative substitutions may give the desired biological activity see Cain SA, Williams DM, Harris V, Monk PN. Selection of novel ligands from a whole-molecule randomly mutated C5a library. Protein Eng. 2001 Mar;14(3):189-93, which is incorporated by reference.

A functionally equivalent polypeptide according to the invention is a variant wherein one or more amino acid residues are substituted with conserved or non-conserved amino acid residues, or one in which one or more amino acid residues includes a substituent group. Conservative substitutions are the replacements, one for another, among the

aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among aromatic residues Phe and Tyr.

5    In addition, the invention features polypeptide sequences having at least 75% identity with the polypeptide sequences as hereindisclosed, or fragments and functionally equivalent polypeptides thereof.    In one embodiment, the polypeptides have at least 85% identity, more preferably at least 90% identity, even more preferably at least 95% identity, still more preferably at least 97% identity, and most preferably at least 99%

10   identity with the amino acid sequences illustrated herein.

In a preferred method of the invention said nucleic acid molecule comprises the nucleic acid sequence encoding the amino acid sequence Ciz1 in Fig 2 or variants thereof.  In a further preferred method of the invention said nucleic acid molecule consists of the

15   nucleic acid sequence which encodes the amino acid sequence Ciz1 in Fig 2 or variants thereof.

In a further preferred method of the invention said polypeptide molecule comprises the amino acid sequence Ciz1 in Fig 2 or variants thereof.  In a further preferred method of

20   the invention said polypeptide molecule consists of the amino acid sequence Ciz1 in Fig 2 or variants thereof.

In a further preferred method of the invention said polypeptide is expressed by a cell, preferably a mammalian cell, or animal and said screening method is a cell-based

25   screening method.    Preferably said cell naturally expresses the Ciz polypeptide. Alternatively said cell is transfected with a nucleic acid molecule encoding a polypeptide with Ciz activity (or a variant molecule found in cancer cells).

10

According to a further aspect of the invention there is provided an agent obtainable by the method according to the invention.

Preferably said agent is an antagonist of Ciz1 mediated DNA replication. Alternatively said agent is an agonist of Ciz1 mediated DNA replication.

In a further preferred method of the invention said agent is selected from the group consisting of: polypeptide; peptide; aptamer; chemical; antibody; nucleic acid.

Preferably said agent is an anti-sense nucleic acid molecule which binds to and thereby blocks or inactivates the mRNA encoded by any of the nucleic acid sequences in (i) above.

In an alternative embodiment, said agent is an RNAi molecule and comprises two complementary strands of RNA (a sense strand and an antisense strand) annealed to each other to form a double stranded RNA molecule. Preferably the RNAi molecule is derived from the exonic sequence of the Ciz1 gene or from another over-lapping gene.

In one embodiment unspliced mRNA is targetted with with RNAi to inhibit production of the spliced variant. In another the spliced variant mRNA is ablated without affecting the non-variant mRNA.

In a preferred method of the invention said peptide is an oligopeptide. Preferably, said oligopeptide is at least 10 amino acids long. Preferably said oligopeptide is at least 20, 30, 40, 50 amino acids in length.

In a further preferred method of the invention said peptide is a modified peptide.

It will be apparent to one skilled in the art that modified amino acids include, by way of example and not by way of limitation, 4-hydroxyproline, 5-hydroxylysine, $N^6$-

acetyllysine, $N^6$-methyllysine, $N^6,N^6$-dimethyllysine, $N^6,N^6,N^6$-trimethyllysine, cyclohexyalanine, D-amino acids, ornithine. Other modifications include amino acids with a $C_2$, $C_3$ or $C_4$ alkyl R group optionally substituted by 1, 2 or 3 substituents selected from halo ( eg F, Br, I), hydroxy or $C_1$-$C_4$ alkoxy.

5

Alternatively said peptide is modified by acetylation and/or amidation.

In a preferred method of the invention the polypeptides or peptides are modified by cyclisation. Cyclisation is known in the art, (see Scott et al Chem Biol (2001), 8:801-

10    815; Gellerman et al J. Peptide Res (2001), 57: 277-291; Dutta et al J. Peptide Res (2000), 8: 398-412; Ngoka and Gross J Amer Soc Mass Spec (1999), 10:360-363).

According to a further aspect of the invention there is provided a vector as a delivery means for antisense or an RNAi molecule which inhibits Ciz1 or variants thereof and

15    thereby allows the targetting of cells expressing the truncated protein.

Preferably the vector includes an expression cassette comprising the nucleotide sequence selected from the group consisting of;

a) the nucleic acid sequence which encodes Ciz1 amino acid sequence as shown

20    in Fig 8a or 8b;

b)a nucleic acid molecule which hybridizes to the nucleic acid sequence of (a) ;

c)a nucleic acid molecule which has a nucleic acid sequence which is degenerate because of the genetic code to the sequences in a) and b) and any sequence which is complimentary to any of the above sequences;

25    d) a nucleic acid sequence that encodes Ciz1 pre-mRNA (i.e., the genomic sequence)

wherein the expression cassette is transciptionally linked to a promoter sequence.

Preferably the vectors including the expression cassette is adapted for eukaryotic gene

30    expression. Typically said adaptation includes, by example and not by way of

12

limitation, the provision of transcription control sequences (promoter sequences) which mediate cell/tissue specific expression. These promoter sequences may be cell/tissue specific, inducible or constitutive.

5    Promoter elements typically also include so called TATA box and RNA polymerase initiation selection sequences which function to select a site of transcription initiation. These sequences also bind polypeptides which function, *inter alia*, to facilitate transcription initiation selection by RNA polymerase.

10    Adaptations also include the provision of selectable markers and autonomous replication sequences which both facilitate the maintenance of said vector in either the eukaryotic cell or prokaryotic host. Vectors which are maintained autonomously are referred to as episomal vectors. Further adaptations which facilitate the expression of vector encoded genes include the provision of transcription termination sequences.

15

These adaptations are well known in the art. There is a significant amount of published literature with respect to expression vector construction and recombinant DNA techniques in general. Please see, Sambrook et al (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbour Laboratory, Cold Spring Harbour, NY and

20    references therein; Marston, F (1987) DNA Cloning Techniques: A Practical Approach Vol III IRL Press, Oxford UK; DNA Cloning: F M Ausubel et al, Current Protocols in Molecular Biology, John Wiley & Sons, Inc.(1994).

According to the present invention there is provided a diagnostic method for the

25    identification of proliferative disorders comprising detecting the expression of the Ciz 1 gene and mutations in the genomic sequence.

Preferably said diagnostic method comprises the steps of :

(i)     contacting a sample isolated from a subject to be tested with an agent which specifically binds a polypeptide with Ciz 1 activity or a nucleic acid molecule encoding a polypeptide with Ciz 1 activity; and

(ii)    detecting or measuring the binding of the agent on said polypeptide or nucleic

5      acid in said sample.

In one embodiment, the diagnostic method of the present invention is carried out *in-vivo*.

Preferably the diagnostic method provides for a quantitative measure of Ciz1 RNA or

10    protein variants in a sample.

In one embodiment of the invention there is provided the use of an agent which modulates Ciz1 RNA or protein, or variants thereof, as a pharmaceutical.

15    Preferably said pharmaceutical comprises an agent identifed by the screening method of the present invention and a pharmacetically acceptable carrier, excipient or diluent.

Preferably said pharmaceutical is for oral or topical administration or for administration by injection.

20

In a further preferred embodiment of the invention there is provided the use of an agent according to the invention for the manufacture of a medicament for use in the treatment of proliferative disease. Preferably said proliferative disease is cancer.

25    Preferably said cancer is a paediatric cancer and is selected from the group consisting of; retinoblastoma, neuroblastoma, Burkitt lymphoma, medulloblastoma.

In an alternate embodiment the disease is liver cancer or metastasis.

14

According to a further aspect of the invention there is provided a method to treat a proliferatvie disease comprising administering to an animal, preferably a human, an agent obtainable by the method according to the invention.

5   Preferably said proliferative disease is cancer and is preferably a paediatric cancer selected from the group consisting of ; retinoblastoma, neuroblastoma, Burkitt lymphoma, medulloblastoma.

In an alternative embodiment the disease is liver cancer or metastasis.

10

According to an alternate aspect of the invention, there is provided the use of an agent according to the invention for the manufacture of a medicament to slow cell division or growth.

15   The invention also includes the use of the Ciz1 amino acid sequence and protein structure in rational drug design and the use of Ciz1 or its variants or derived peptides thereof for screening chemical libraries for agents that specifically bind to Ciz1.

An embodiment of the invention will now be described by example only and with
20   reference to the following figures:

Fig. 1 Illustrates the effect of cyclin A-cdk2 on late G1 nuclei. A) DNA synthesis is activated by recombinant cyclin A-cdk2. Late G1 nuclei (harvested 17 hours after release from quiescence) were incubated in mid-G1 phase (15 hour) cytosolic extract
25   supplemented with recombinant cyclin-dependent kinases at the indicated concentrations. The number of replicating nuclei increased in the presence of cyclin A-cdk2 (but not cyclin E-cdk2), within a narrow range of concentrations. Control bar shows the number of nuclei in this population that are already in S phase (unshaded) and the fraction that are induced to replicate by S phase extract (shaded). B) Detection of
30   mouse p85. Asynchronous 3T3 cells were separated into soluble and insoluble fractions

by hypotonic lysis and centrifugation. Anti-human Cdc6 antibody V1 reacts with mouse Cdc6 and a second antigen in the 85kDa range, that is present in both fractions. C) Late G1 nuclei (17 hour) were incubated in mid-G1 extract in the presence of recombinant cyclin A-cdk2 (as indicated). After 15 minutes nuclei were washed and the chromatin

5    fraction was isolated and separated by SDS-Page. P85 accumulates in the chromatin fraction in the presence of cyclin A-cdk2, peaking at the same concentration as initiation of DNA replication. Chromatin bound Mcm3 is shown as a control.

Fig. 2 Illustrates the embryonic form mouse Ciz1 protein aligned with the predicted full-

10   length form. Embryonic form mouse Ciz1 (ECiz1) lacks three sequence blocks compared to the predicted full-length form. Blue stars indicate amino acids which have been targeted by site-directed mutagenesis.

Fig. 3 Illustrates Ciz1 splice variants. A) Summary of mouse Ciz1 variants. In addition

15   to the three variant sequence blocks (encoded by exons 1 and 2, and part of exons 6 and 8) that were identified by analysis of ECiz1, Ciz1 from mouse ES cells lacks a fourth region which corresponds to exons 3 and 4. Grey bar represents unchecked and potentially variable sequence. B) Human Ciz1 is also alternately spliced with variability occurring in the same four regions. Human Ciz1 variants are all derived from pediatric

20   cancers or from embryonic source. Grey bar represents uncertain sequence that may be derived from an alternate reading frame. C) Similarity between mouse and human variable exons. Identical amino acids are shown in red.

Fig. 4 Illustrates ECiz1 promotes initiation of mammalian DNA replication. A) *In vitro*

25   replication reactions containing late G1 nuclei and S phase cytosolic extract were supplemented with purified recombinant ECiz1. Histogram shows the average number of replicating nuclei with and without ECiz1, and standard deviations. *In vitro* initiating nuclei (black) are shown above the 'background' level of S phase nuclei in this population (unshaded). Images show nuclei incubated in the absence (i) or presence (ii)

30   of ECiz1 (yellow). Total nuclei are counterstained with propidium iodode (red). B) The

16

response to recombinant ECiz1 is concentration dependent with a sharp optimum in the nM range. In this experiment, and those shown in figs 4C and D, and in figs 6A, B and C, results are expressed as % initiation rather than % replication. This is calculated from the number of nuclei that initiate *in vitro* and the number of nuclei that are 'competent'

5    to initiate *in vitro* (see methods). C) The replication function of ECiz1 resides in the N-terminal fragment, N-term442 (shown in fig. 5B). D) No initiation activity is detected in the C-terminal fragment, C-term274 (shown in fig. 5B), when assayed over a broad range of concentrations.


10   Fig. 5 Illustrates Mouse Ciz1 protein and derived expression constructs A) Embryonic mouse Ciz1 protein showing sequence features and putative domains. Consensus cdk phosphorylation sites that are conserved (white stars) or not conserved (black stars) in human Ciz1 are indicated. The region containing the putative p21$^{cip1}$ interacting sequence (white dots) is defined in mouse embryonic Ciz1 by comparison with a

15   truncated human Ciz1 construct that interacts with p21 $^{cip1}$ (Mitsui et al., 1999). The positions of sequences that are absent in embryonic form Ciz1 are indicated by white triangles. B) Embryonic Ciz1 (ECiz1) and derived constructs used to produce protein for replication assays. Numbers in parentheses relate to amino-acid positions in the predicted full-length form of Ciz1 shown in fig. 2.

20
Fig. 6 illustrates Ciz1 is regulated by cyclin-dependent kinases. A) Stimulation of initiation by recombinant ECiz1 mutant T(191/2)A, in which two threonines in the conserved consensus cdk phoshporylation site at amino acids 191 and 192 have been changed to alanines. Unlike ECiz1, this protein is not subject to down-regulation of

25   activity at high concentrations indicating that endogenous cdk's regulate Ciz1 activity.


Figure 7 Endogenous Ciz1 is present at the same sites in the nucleus where DNA replication takes place.   Recombinant Eciz1 was used to raise a specific rabbit polyclonal antibody that recognises endogenous human and mouse Ciz1. The antibody

30   was applied to asynchronously growing mouse 3T3 cells and to human HeLa cells after

detergent extraction (to remove soluble proteins) and fixation. In both cell types anti-ECiz1 reveals a punctate staining pattern that is similar to that seen with antibodies to DNA replication proteins. A) Dual detection of Ciz1 (red), and the DNA replication factor PCNA with monoclonal antibody PC10 (Sigma, green) reveals that almost all

5  PCNA staining overlaps with Ciz1 staining in the merged image, but that additional regions of Ciz1 staining exist. An S phase cell that contains detergent resistant PCNA is shown. B) A monoclonal anti-human SC35 antibody (Sigma) reveals the sites at which RNA splicing takes place in the nucleus of HeLa cells (green). Human Ciz1 does not co-localize with SC35 in these cells. All methods are as previously described (Coverley et

10  al 2001 or references therein).


Figure 8 illustrates a) mouse full length cDNA sequence b)human full length cDNA sequence c) human full length protein sequence.


15  **Materials and Methods**

Cloning. A lamba triplEx  5'-stretch, full length enriched cDNA expression library derived from 11 Day old mouse embryos (Clonetech ML5015t) was used to infect *E. coli* X11blue according to the recommended protocol (Clonetech). Plaques were lifted onto 0.45 micron nitrocellulose filters pre-soaked in 10mM IPTG (Sigma). Affinity

20  purified antibody V1 was applied to approximately $3 \times 10^6$ plaques at 1/1000 dilution in PBS, 10% non-fat milk powder, 0.4% Tween20, after blocking for 30 minutes in the absence of antibody. After two hours filters were washed three times with the same buffer and reactive plaques were visualized with anti-rabbit secondary antibody conjugated to horse-radish peroxidase (Sigma), and enhanced chemi-luminescence

25  (ECL, Amersham) according to standard procedures. 43 independent plaques were picked but only two strains of phage survived a further three rounds of screening. These were converted to pTriplEx by transforming into BM25.8 and sequenced. One codes for mouse Cdc6 (clone P) and the other (clone L) for an unknown mouse protein that is homologous to human Ciz1.   Sequence alignments were performed with Multialin

30  (Corpet, 1998).

<u>Antibodies</u>. Rabbit polyclonal antibody V1 was raised against an internal fragment of bacterial expressed human Cdc6 corresponding to amino-acids 195-412 and affinity purified against this fragment by standard procedures (Harlow and Lane, 1988).

5

<u>Constructs.</u> pGEX-ECiz1 was generated by inserting a 2.3kb SmaI-XbaI (blunt ended) fragment from pTriplEx-clone L into the SmaI site of pGex-6P-3 (Amersham). pGEX-Nterm442 was generated by inserting the 1.35kb XmaI-XhoI fragment into XmaI-XhoI digested pGex-6P-3, and pGEX-Cterm274 by inserting the 0.95kb XhoI fragment into

10   XhoI digested pGex-6P-3.

pGEX-T(191/2)A was generated from pGEX-ECiz1 by site directed mutagenesis (Stratagene Quikchange 200518-5) of the conserved consensus cdk phosphorylation site at amino-acids 191 and 192 using primers and

15   AACCCCCTCTTCCGCCGCCCCCAATCGCAAGA
TCTTGCGATTGGGGGCGGCGGAAGAGGGGGTT, producing two alanines in place of two threonines. pGEX-T(293)A was generated from pGEX-ECiz1 by site directed mutagenesis using primers AAGCAGACACAGGCCCCGGATCGGCTGCCT and AGGCAGCCGA TCCGGGGCCTGTGTCTGCTT changing the threonine at 293 to

20   alanine.
All clones were verified by DNA sequence analysis, prior to transfer into *E. coli* BL21 for protein expression.

<u>Protein expression</u> Recombinant Ciz1, Ciz1 fragments and point mutants were produced
25   in BL21-pLysS (Stratagene) as glutathione S-transferase-tagged protein. This was purified from sonicated and cleared bacterial lysates by binding to glutathione sepharose 4B (Amersham). Recombinant protein was eluted by cleavage from the GST tag using precision protease (as recommended by the manufacturer, Amersham), into buffer (50mM Tris-HC pH 7.0, 150mM NaCl, 1mM DTT). This yielded protein preparations
30   between 0.2 and 2.0 mg/ml. For replication assays serial dilutions were made in 100mM

Hepes pH 7.8, 1mM DTT, 50% glycerol so that not more than 1μl of protein solution was added to 10μl replication assays, yielding the concentrations shown.

Cell synchrony Mouse 3T3 cells were synchronized by release from quiescence as previously described (Coverley et al., 2002). Nuclei were prepared from cells harvested between 16 and 18 hours after release, yielding population containing S phase nuclei, replication competent late G1 nuclei and unresponsive early G1/G0 nuclei, in varying proportions. Recipient, mid-G1 3T3 extracts were prepared from cells harvested 15 hours after release from quiescence (these typically contain approximately 5% S phase cells). S phase extracts were prepared from HeLa cells released for two hours from two sequential thymidine-induced S phase blocks, as outlined (Krude et al., 1997). HeLa cells are used for S phase extracts because they are easily synchronized in large quantities.

Replication assays Nuclei and extract preparation has been previously described (Coverley et al., 2000; Krude et al., 1997; Stoeber et al., 1998). *In vitro* replication assayed were performed as described (Coverley et al., 2002). Reactions containing 10μl of extract (supplemented with energy regenerating system and nucleotides including biotinylated dUTP), and $5 \times 10^4$ nuclei were incubated with Ciz1 at the indicated concentrations, for 60mins at 37°C. Reactions were stopped by the addition of 50μl of 0.5% triton X100 and fixed by the addition of 50 μl of 8% paraformaldehyde. After transfer to coverslips nuclei were stained with streptavidin-FITC (Amersham) to reveal DNA synthesized *in vitro*, and counterstained with Toto-3-iodide. The proportion of labeled nuclei in each sample was quantified by inspection at 1000X magnification. Nuclei with fluorescent foci were scored positive. Images of *in vitro* replicating nuclei were generated by confocal microscopy at 600X magnifications, of samples counterstained with propidium iodide.

Data analysis and presentation Prior to use in initiation assays each preparation of synchronized G1 phase nuclei is tested so that the proportion of nuclei that are already in

S phase is established ('%S'). To do this nuclei are incubated in an extract that is incapable of inducing initiation of DNA synthesis (from mid-G1 phase cells harvested 15 hours after release from quiescence), but that will efficiently support elongation DNA synthesis from origins that were initiated *in vivo*. The elongating fraction of nuclei

5      incorporates labeled nucleotides efficiently during *in vitro* initiation assays but is uninformative. Routinely this fraction is pre-established and subtracted from the raw data. Synchronized populations in which 20% or less are in S phase are used for initiation assays.

10     Similarly, for each preparation of nuclei the fraction that is capable of responding to inducers of S phase and initiating DNA synthesis *in vitro* (the competent fraction) is also pre-established. This is taken to be the maximum proportion of nuclei that replicate *in vitro* under any conditions, and is usually around 50% of the total population.

15     When 3T3 cells are released from quiescence by the protocol used here a maximum of 70% of the total population enters S phase *in vivo* (Coverley et al., 2002) while the rest do not re-enter the cell cycle. This defines the maximum possible initiation frequency at around 70%. For most populations this is nearer 60%, so the proportion of G1 nuclei that replicate *in vitro* (around 50%) is comparable to the proportion that would have entered

20     S phase *in vivo*.

The late G1 population of 3T3 nuclei used in the replication experiments shown here was harvested 17 hours after release from quiescence. 17% of these were already in S phase (fig. 4A) and the maximum number that have replicated in any assay *in vitro* is

25     50%. Therefore, 33% of this total population is in late G1 and competent to replicate (%C). For each data point the initiating fraction ('% initiation') is calculated using %S and %C according to the formula, % Initiation = (% replication -%S)/%C.

**Results**

Identification of Ciz 1Late G1 3T3 nuclei (isolated 17 hours after release from quiescence) can be induced to begin DNA synthesis by recombinant cyclin A-cdk2

5    ((Coverley et al., 2002) and fig. 1A). The response to cyclin A-cdk2 is strictly dependent upon the concentration of active kinase that is added to the recipient extract (from mid-G1 phase cells harvested 15 hours after release from quiescence). When nuclei are incubated under these conditions a number of changes to chromatin bound proteins can be seen. We have identified of one of these proteins by expression cloning.

10

Our previous work on replication complex assembly proteins generated a polyclonal antibody against human Cdc6 (V1), which has been used in previous studies (Coverley et al., 2000; Stoeber et al., 1998; Williams et al., 1998). When applied to material derived from mouse 3T3 cells antibody V1 reacts with Cdc6 and another protein of

15    approximately 85kDa (fig. 1B). Antibody V1 has been used to follow p85 through *in vitro* replication complex assembly and activation reactions. Strikingly, p85 fractionates with chromatin at the same cyclin A-cdk2 concentration that activates DNA synthesis *in vitro* and becomes resistant to extraction (fig. 1C). This implicates p85 in initiation of DNA replication.

20

When applied to a cDNA expression library derived from 11-day mouse embryos antibody V1 picked out two clones that survived multiple rounds of screening. One encodes mouse Cdc6, while the other encodes 716 amino acids of an uncharacterized mouse protein, homologous to human Ciz1. Human Ciz1 (Accession Number

25    AB030835), the predicted mouse Ciz1 cDNA sequence (XM123748) and a full-length cDNA clone derived from a mouse mammary tumour library (No. BC018483) all contain three additional sequence blocks that are not present in our embryonic mouse Ciz1 clone (fig. 2A). These correspond to exons 1 and 2, and parts of exons 6 and 8 of the mouse gene. Our analysis has focused on the activity that resides in the embryonic

30    form of mouse Ciz1 (ECiz1) that lacks these exons (fig. 2).

22

**Human tumours express Ciz1 variants that resemble mouse embryonic Ciz1** Human Ciz1 also exists as multiple isoforms, the functions of which are unstudied. The cDNA encoding the predicted full-length form (XM 026951) was isolated from a B-cell library (Mitsui et al., 1999), while a number of variant sequences have been found by large-scale genome analysis projects (HGMP, IMAGE, NCBI, HEMBA). Ciz1 variants (AK023978, BC004119, BC021163, AF234161, AK027287) lack one or more of the same exons that are absent in mouse embryonic Ciz1. Furthermore, they are all derived from paediatric cancers (retinoblastoma, neuroblastoma, Burkitt lymphoma, medulloblastoma) or from a human embryo. The data on Ciz1 isoforms has been collated and is summarised in fig.3.

**Ciz1 stimulates initiation *in vitro*** Recombinant ECiz1 protein stimulates initiation of DNA replication in late G1 3T3 nuclei (18 hour), when used to supplement extracts from S phase HeLa cells (fig. 4A). Upon addition of Eciz1 the number of nuclei that replicated *in vitro* was increased from 30%(+/-0.856) to 46% (+/- 5.52). This increase is in addition to the 13% of nuclei that initiate replication in response to S phase cytosol alone, and the 17% that are already in S phase. When combined with the observation that p85-Ciz1 associates with chromatin under the same conditions that support initiation *in vitro*, this result argues strongly for a role for Ciz1 in initiation of mammalian DNA replication.

Stimulation of initiation is sensitive to the concentration of recombinant ECiz1, with peak activity at around 2 nM (fig. 4B). This concentration dependent response to ECiz1 echoes the results of previous experiments in which recombinant cdks were added to mammalian cell-free replication systems (Coverley et al., 2002; Krude et al., 1997) and fig. 1A). In all of these experiments stimulation of initiation is lost when an excess of active protein is added.

Cizl is a multi-functional protein ECiz1 possesses several sequence features in its C-terminal third that suggest possible functions (fig. 5A). The matrin 3 domain implies interaction with the nuclear matrix and the three C2H2 type zinc-fingers suggest interaction with DNA or RNA. Here, we report that the replication function of ECiz1

5  does not require these domains. N-term442 (fig. 5B), which lacks the matrin three domain and two of the zinc fingers, stimulates initiation in the same way as intact ECiz1 (fig. 4C). The proportional increase in initiation and the active concentration are essentially the same for both forms of the protein. Furthermore, the C-terminal portion, C-term274 contains no residual replication activity in our assay (fig. 4D). These data

10  show that, when assayed in trans, the nucleotide and matrix interaction domains do not contribute to ECiz1 replication function. It remains possible however, that these domains play a role in localizing endogenous Cizl to specific sites in the nucleus *in vivo*.

ECiz1 replication function is regulated by cyclin dependent kinases. ECiz1 contains five
15  sequences that conform to the consensus for cdk phosphorylation, suggesting that cyclin dependent kinases regulate Ciz1 function (fig. 5A). Two of these sites are conserved in human and mouse Ciz1. One of these is in the C-terminal portion of Ciz1, which is not required for initiation function, while the remaining site is located in the N-terminal portion adjacent to the site at which exon 6 is alternately spliced. We have mutated this
20  cdk site in ECiz1, changing two threonines at 191 and 192 to two alanines, generating ECiz1T(191/2)A. Like ECiz1, recombinant ECiz1T(191/2)A stimulates initiation of DNA replication in late G1 nuclei (fig. 6A). It does this to a similar extent and at a similar concentration as ECiz1, however unlike ECiz1 the response to ECiz1T(191/2)A is not down regulated when it is added at high concentrations. This demonstrates that the
25  stimulatory effect of ECiz1 on initiation of DNA replication is regulated by cyclin-dependent kinases. Furthermore, it suggests a sensing mechanism that monitors the level of Ciz1, which is capable of distinguishing functional levels (nM range) from excess (10 nM range).

The downturn in the response curve for ECiz1 (fig. 4B) returns initiation to the level achieved with S phase extract alone (in the absence of added Ciz1), rather than to the level achieved with G1 phase extract (which only supports elongation DNA synthesis). Therefore, negative regulation of ECiz1 dependent initiation does not affect all initiation events, only those influenced by recombinant ECiz1.

The consensus cdk phosphorylation site at amino acid 293 was also mutated with no affect on replication activity (fig. 6C).

Ciz1 colocalizes with the replication factor PCNA, but not the splicing factor SC35, indicating that Ciz1 is present at the sites where DNA replication takes place (See Figure 7).

**Discussion**

Embryonic form Ciz1 (Eciz1) positively regulates initiation of mammalian DNA replication and is found at site of DNA replication inside the nucleus of cells. Its chromatin binding and replication function is regulated by cdk2. It may also modulate cdk2 activity via p21. This demonstration of a positive role in DNA replication initiation combined with the embryo and cancer-derived splice variants suggest the following working hypothesis.

Replication activity resides in the embryonic form of Ciz1 (proven). Inclusion of additional exons by a regulated splicing mechanism during the course of normal development, adds functions that are not present in the embryonic form. This could selectively restrain the initiation function of Ciz1 in response to external signals. Expression of embryonic forms of Ciz1 at the wrong time in development could lead to cancer.

For example, one of the variable exons encodes a short conserved DSSSQ sequence motif that is absent in mouse ECiz1 and in a human medulloblastoma. This is directly

adjacent to the consensus cdk phosphorylation site that we have shown to be involved in regulation of ECiz1 function. Conditional inclusion of the DSSSQ sequence might make Ciz1 the subject of regulation by the ATM/ATR family of protein kinases, which phosphorylate proteins at SQ sequences, thereby restraining Ciz1 initiation function in response to DNA damage.

In summary, variable exon usage, amino acid sequence and overall gene structure are highly conserved between mouse and human Ciz1, and forms derived from early in development (or from cancer cells) have fewer exons than differentiated cells. The functions of these variable exons are unstudied and it is possible that errors in Ciz1 splicing during development could contribute to the formation or progression of cancer cell lineages by allowing Ciz1 to escape regulatory pathways and promote DNA replication at the wrong time.

**References**

Blow, J. J. (2001). Control of chromosome DNA replication in the early *Xenopus* embryo. EMBO J *20*, 3293-3297.

Corpet, F. (1998). Multiple sequence alignment with hierarchical clustering. Nucl. Acids Res. *16*, 10881-10890.

Coverley, D., Laman, H., and Laskey, R. A. (2002). Distinct roles for cyclins E and A during DNA replication complex assembly and activation. Nat Cell Biol. *4*, 523-528.

Coverley, D., Pelizon, C., Trewick, S., and Laskey, R. A. (2000). Chromatin bound Cdc6 persists in S and G2 phases in human cells, while soluble Cdc6 is destroyed in a cyclin A-cdk2 dependent process. J Cell Sci. *113*, 1929-1938.

Diffley, J. F. X., and Labib, K. (2002). The chromosome replication cycle. J Cell Sci *115*, 869-872.

Harlow, E., and Lane, D. (1988). Antibodies : A laboratory manual (New York, Cold Spring Harbour laboratory press).

Krude, T. (2000). Initiation of human DNA replication *in vitro* using nuclei from cells arrested at an initiation-competent state. J Biol. Chem. *275*, 13699-13707.

Krude, T., Jackman, M., Pines, J., and Laskey, R. A. (1997). Cyclin/Cdk-dependent initiation of DNA replication in a human cell-free system. Cell *88*, 109-119.

5    Laman, H., Coverley, D., Krude, T. K., Laskey, R. A., and Jones, N. (2001). Viral cyclin/cdk6 complexes initiate nuclear DNA replication. Mol Cell Biol *2*, 624-635.

Mitsui, K., Matsumoto, A., Ohtsuka, S., Ohtsubo, M., and Yoshimura, A. (1999). Cloning and Characterization of a Novel p21cip1/waf1-interacting Zinc Finger Protein, Ciz1. Biochem. Biophys. Res. Com. *264*, 457-464.

10    Stoeber, K., Mills, A. D., Kubota, Y., Krude, T., Romanowski, P., Marheineke, K., Laskey, R. A., and Williams, G. H. (1998). Cdc6 protein causes premature entry into S phase in a mammalian cell-free system. EMBO J *17*, 7219-7229.

Williams, G. H., Romanowski, P., Morris, L., Madine, M., Mills, A. D., Stoeber, K., Marr, J., Laskey, R. A., and Coleman, N. (1998). Improved cervical smear assessment
15    using antibodies against proteins that regulate DNA replication. Proc Natl. Acad. Sci. USA *95*, 14932-14937.

# Fig. 1

## Cyclin A-cdk2 dependent initiation of DNA replication and chromatin binding of p85

# Fig. 2
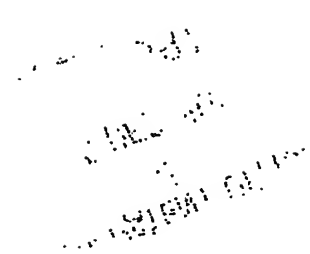
**Full length and embryonic form mouse Ciz1**

# Fig. 3

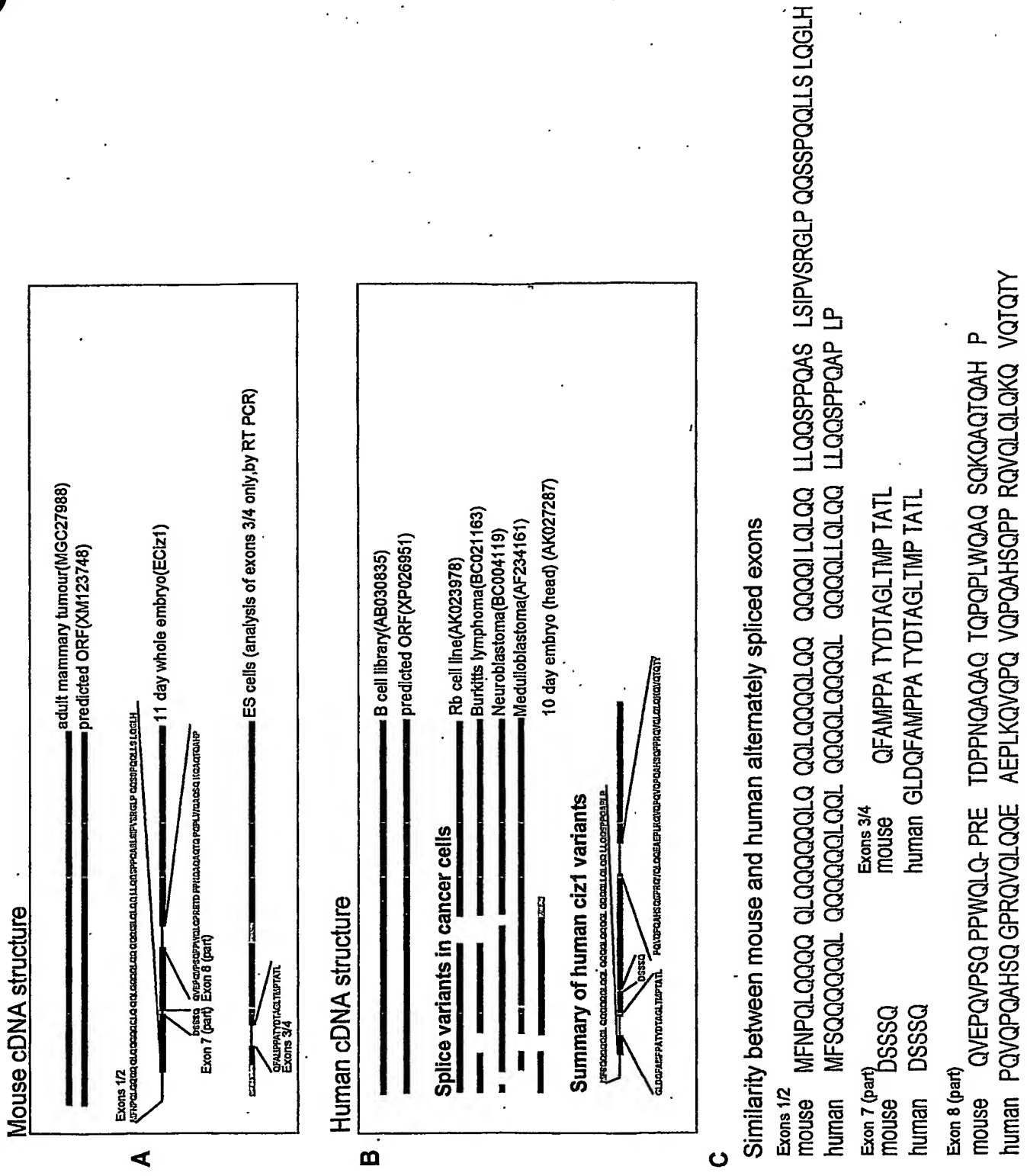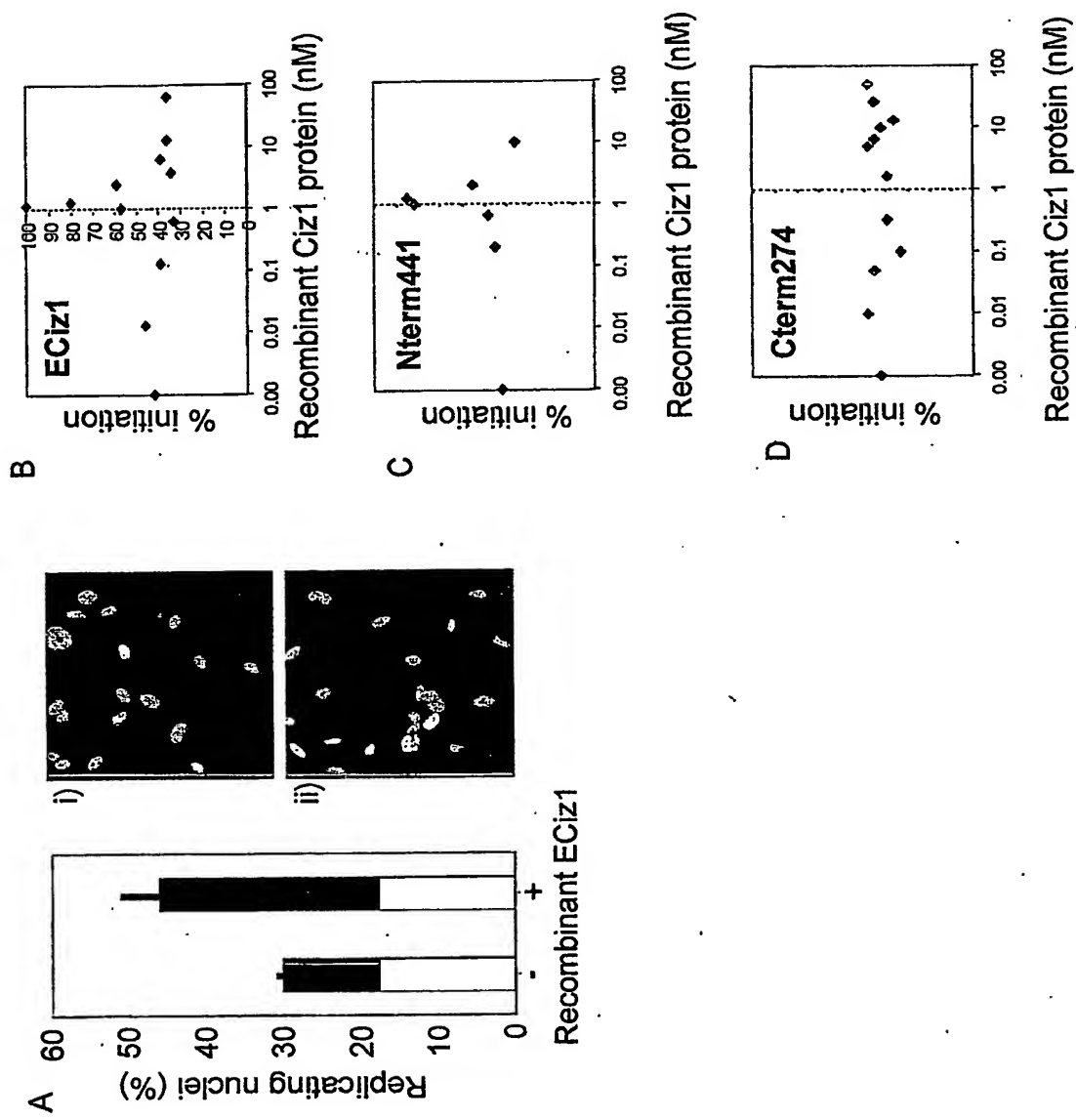Ciz1 is alternately spliced in mouse and human embryos and may be mis-spliced in pediatric cancers

**A** Mouse cDNA structure

adult mammary tumour(MGC27988)
predicted ORF(XM123748)

11 day whole embryo(ECiz1)

ES cells (analysis of exons 3/4 only,by RT PCR)

Exons 1/2
Exon 7 (part) Exon 8 (part)
Exons 3/4

**B** Human cDNA structure

B cell library(AB030835)
predicted ORF(XP026951)

Splice variants in cancer cells

Rb cell line(AK023978)
Burkitts lymphoma(BC021163)
Neuroblastoma(BC004119)
Medulloblastoma(AF234161)

10 day embryo (head) (AK027287)

Summary of human ciz1 variants

**C** Similarity between mouse and human alternately spliced exons

Exons 1/2
mouse   MFNPQLQQQQ QLQQQQQQLQ QQLQQQQLQQ QQQQILQLQQ LLQQSPPQAS  LSIPVSRGLP QQSSPQQLLS LQGLH
human   MFSQQQQQQL QQQQQQLQQL QQQLQQQL  QQQQLLQLQQ LLQQSPPQAP  LP

Exon 7 (part)                    Exons 3/4
mouse   DSSSQ            mouse   QFAMPPA TYDTAGLTMP TATL
human   DSSSQ            human   GLDQFAMPPA TYDTAGLTMP TATL

Exon 8 (part)
mouse   QVEPQVPSQ PPWQLQ- PRE   TDPPNQAQAQ TQPQPLWQAQ SQKQAQTQAH  P
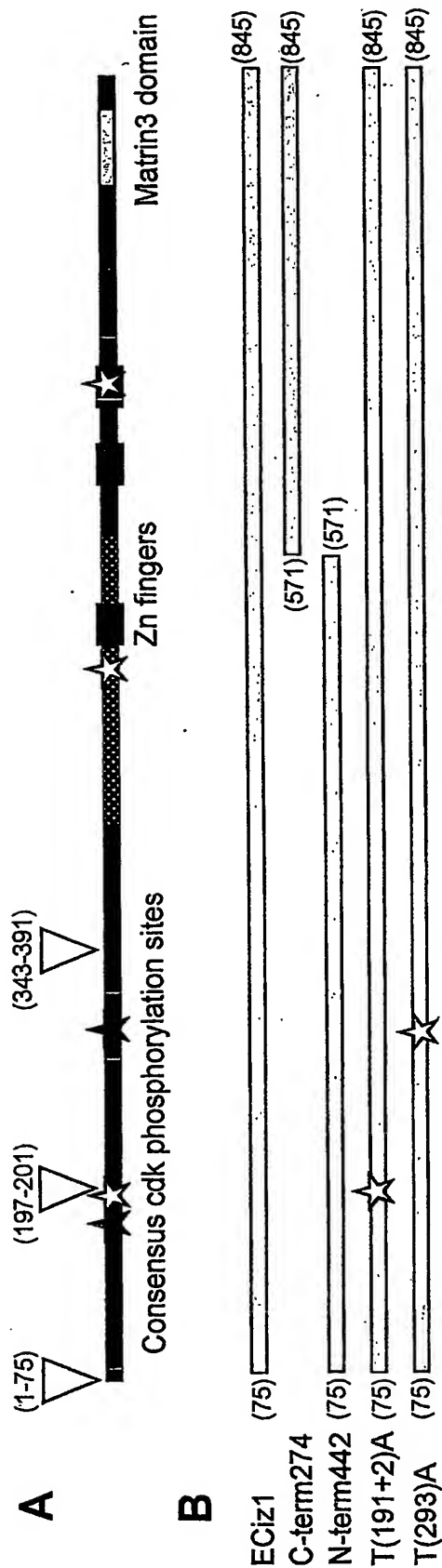human   PQVQPQAHSQ GPRQVQLQQE AEPLKQVQPQ VQPQAHSQPP RQVQLQLQKQ  VQTQTY

# Fig. 4

## Recombinant Ciz1 promotes initiation of mammalian DNA replication *in vitro*

# Fig. 5

**Embryonic Ciz1 protein structure and expression constructs**



A

(1-75)  (197-201)  (343-391)

Consensus cdk phosphorylation sites

Zn fingers

Matrin3 domain

B

| | | |
|---|---|---|
| ECiz1 | (75) | (845) |
| C-term274 | | (845) |
| N-term442 | (571) | (571) |
| T(191+2)A | (75) | (845) |
| T(293)A | (75) | (845) |

## FIGURE 6
## Ciz1 replication function is negatively regulated by cyclin-dependent kinase

A    T(191/2)A



Recombinant Ciz1 protein (nM)

A

S phase 3T3

Ciz1    PCNA    merge

B

AS HeLa cells

Ciz1    SC35    merge
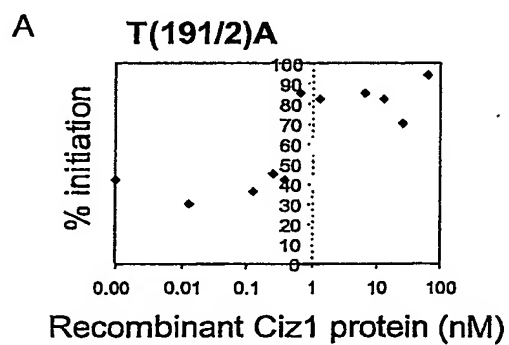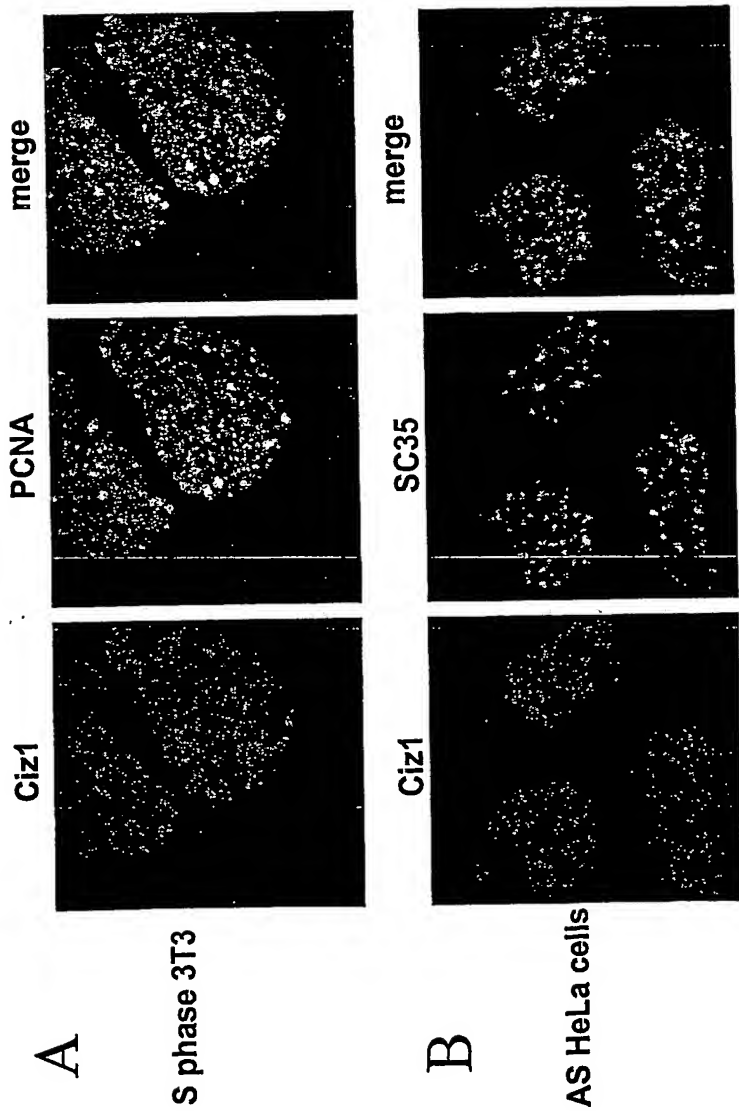
Figure 7.

## Figure 8a; Mouse full length cDNA sequence

CATGTTCAAC CCGCAACTCC AGCAGCAGCA ACAGTTGCAG CAGCAGCAGC
AACAGTTGCA GCAGCAGCTC CAGCAGCAGC AGCTCCAGCA GCAGCAACAG
CAGATACTGC AGCTCCAACA GCTGCTGCAA CAGTCCCCAC CACAGGCCTC
CTTGTCCATT CCTGTCAGCC GGGGCCTCCC CCAGCAGTCA TCCCCGCAAC
AGCTTCTGAG TCTCCAGGGC CTCCACTCGA CCTCCCTGCT CAATGGCCCC
ATGCTGCAAA GAGCTTTGCT CCTACAGCAG TTGCAAGGAC TGGACCAGTT
TGCAATGCCA CCAGCCACGT ATGACGGTGC CAGCCTCACC ATGCCTACGG
CAACACTGGG TAACCTCCGT GCTTTCAATG TGACAGCCCC AAGCCTAGCA
GCTCCCAGCC TTACACCACC CCAGATGGTC ACCCCAAATC TGCAGCAGTT
CTTTCCCCAG GCTACTCGAC AGTCTCTGCT GGGGCCTCCT CCTGTTGGGG
TCCCAATAAA CCCTTCTCAG CTCAACCACT CAGGGAGGAA CACCCAGAAA
CAGGCCAGAA CCCCCTCTTC CACCACCCCC AATCGCAAGG ATTCTTCTTC
TCAGACGGTG CCTCTGGAAG ACAGGGAAGA CCCCACAGAG GGGTCTGAGG
.AAGCCACGGA GCTCCAGATG GACACATGTG AAGACCAAGA TTCACTAGTC
GGTCCAGATA GCATGCTGAG TGAGCCCCAA GTGCCTGAGC CTGAGCCCTT
TGAGACATTG GAACCACCAG CCAAGAGGTG CAGGAGCTCA GAGGAGTCCA
CCGAGAAAGG CCCTACAGGG CAGCCACAAG CAAGGGTCCA GCCTCAGACC
CAGATGACAG CACCAAAGCA GACACAGACC CCGGATCGGC TGCCTGAGCC
ACCAGAAGTC CAAATGCTGC CGCGTATCCA GCCACAGGCA CTGCAGATCC
AGACCCAGCC AAAGCTGCTG AGGCAGGCAC AGACACAGAC CTCTCCAGAG
CACTTAGCGC CCCAGCAGGA TCAGGTAGAG CCACAGGTAC CATCACAGCC
CCCATGGCAG TTGCAGCCAC GGGAGACAGA CCCACCGAAC CAAGCTCAGG
CACAGACCCA GCCTCAGCCC CTCTGGCAGG CGCAGTCACA GAAGCAGGCC
CAGACACAGG CACATCCACA GGTACCCACC CAAGCACAGT CACAGGAGCA
GACATCAGAG AAGACCCAGG ACCAGCCTCA GACCTGGCCA CAGGGGTCAG
TACCCCCACC AGAACAAGCG TCAGGTCCAG CCTGTGCCAC GGAACCACAG
CTATCCTCTC ACGCTGCAGA AGCTGGGAGT GACCCAGACA AGGCCTTGCC
AGAACCAGTA AGTGCCCAGA GCAGTGAAGA CAGGAGCCGG GAGGCGTCCG
CTGGTGGCCT GGATTTGGGA GAATGTGAAA AGAGAGCGGG AGAGATGCTG
GGGATGTGGG GGGCTGGGAG CTCCCTGAAG GTCACCATCC TGCAGAGTAG
CAACAGCCGG GCCTTTAACA CCACACCCCT CACATCTGGA CCTCGCCCTG
GGGACTCTAC CTCTGCCACC CCTGCCATTG CCAGCACACC CTCCAAGCAA
AGCCTCCAGT TCTTCTGCTA CATCTGCAAG GCCAGCAGCA GCAGCCAGCA
GGAGTTCCAG GATCACATGT CAGAGGCTCA GCACCAACAG CGGCTTGGGG
AAATACAACA CTCGAGCCAG ACCTGCCTGC TGTCCCTGCT GCCCATGCCT
CGGGACATCC TGGAGAAAGA AGCGGAAGAT CCTCCGCCCA AACGCTGGTG
CAACACCTGC CAGGTGTACT ACGTGGGAGA CTTGATCCAG CACCGTAGGA
CACAGGAGCA CAAGGTTGCC AAACAATCCC TGAGGCCCTT CTGCACCATA
TGCAACCGTT ACTTCAAGAC CCCTCGAAAG TTTGTGGAGC ACGTGAAGTC
CCAGGGACAC AAGGACAAGG CCCAAGAGCT GAAGACACTTGAAAAGGAGA
CAGGCAGCCC AGATGAGGAC CACTTCATCA CTGTGGACGC CGTCGGTTGC
TTTGAGAGTG GTCAAGAAGA GGACGAGGAT GACGACGAGGAAGAAGAAGA
AGAAGGAGAG ATTGAGGCTG AGGAGGAATT CTGCAAGCAG GTGAAGCCGA
GAGAAACATC CTCAGAGCAA GGGAAGGGCT CTGAGACGTA CAACCCCAAC
ACAGCCTATG GTGAGGATTT CCTGGTGCCA GTGATGGGCT ATGTCTGTCA

AATCTGTCAC AAGTTCTACG ACAGCAACTC AGAATTGCGG CTTTCTCACT
GCAAGTCCCT GGCCCACTTT GAGAACCTGC AGAAATACAA AGCCAAGAAC
CCAAGCCCTC CTCCTACCCG GCCTGTGAGC CGCAAGTGTG CCATCAACGC
CCGCAACGCC CTGACTGCAC TGTTCACCTC TAGCCACCAG CCCAGCCCCC
AGGACACAGT GAAAATGCCC AGCAAGGTGA AGCCTGGATC CCCCGGACTC
CCTCCTCCCC TTCGGCGCTC AACACGCCTC AAAACCTGAT AGAGGGAGCT
CTGGCCACTC AGCCTGACTA AGGCTCAGTC TGCTAATGCT TCCTAGGTAT
CTGTGTAGAA ATGTTCAAGT GGTTGGTGTT TTTACTCAAA ATCCAATAAA
GAGTCAGTAG TTTGGCAAAA AAAAAAAAA AAAAAAA

## Figure 8b; Human full length cDNA sequence

```
TGGGGGCTGC GGGGCCGGCC CATCCGTGGG GGCGACTTGA GCGTTGAGGG
CGCGCGGGGA GGCGAGCCAC CATGTTCAGC CAGCAGCAGC AGCAGCTCCA
GCAACAGCAG CAGCAGCTCC AGCAGTTACA GCAGCAGCAG CTCCAGCAGC
AGCAATTGCA GCAGCAGCAG TTACTGCAGC TCCAGCAGCT GCTCCAGCAG
TCCCCACCAC AGGCCCCGTT GCCCATGGCT GTCAGCCGGG GGCTCCCCCC
GCAGCAGCCA CAGCAGCCGC TTCTGAATCT CCAGGGCACC AACTCAGCCT
CCCTCCTCAA CGGCTCCATG CTGCAGAGAG CTTTGCTTTT ACAGCAGTTG
CAAGGACTGG ACCAGTTTGC AATGCCACCA GCCACGTATG ACACTGCCGG
TCTCACCATG CCCACAGCAA CACTGGGTAA CCTCCGAGGC TATGGCATGG
CATCCCCAGG CCTCGCAGCC CCAGCCTCA CACCCCACA ACTGGCCACT
CCAAATTTGC AACAGTTCTT TCCCCAGGCC ACTCGCCAGT CCTTGCTGGG
ACCTCCTCCT GTTGGGGTCC CCATGAACCC TTCCCAGTTC AACCTTTCAG
GACGGAACCC CCAGAAACAG GCCCGGACCT CCTCCTCTAC CACCCCCAAT
CGAAAGGATT CTTCTTCTCA GACAATGCCT GTGGAAGACA AGTCAGACCC
CCCAGAGGGG TCTGAGGAAG CCGCAGAGCC CCGGATGGAC ACACCAGAAG
ACCAAGATTT ACCGCCCTGC CCAGAGGACA TCGCCAAGGA AAAACGCACT
CCAGCACCTG AGCCTGAGCC TTGTGAGGCG TCCGAGCTGC CAGCAAAGAG
ATTGAGGAGC TCAGAAGAGC CCACAGAGAA GGAACCTCCA GGGCAGTTAC
AGGTGAAGGC CCAGCCGCAG GCCCGGATGA CAGTACCGAA ACAGACACAG
ACACCAGACC TGCTGCCTGA GGCCCTGGAA GCCCAAGTGC TGCCACGATT
CCAGCCACGG GTCCTGCAGG TCCAGGCCCA GGTGCAGTCA CAGACTCAGC
CGCGGATACC ATCCACAGAC ACCCAGGTGC AGCCAAAGCT GCAGAAGCAG
GCGCAAACAC AGACCTCTCC AGAGCACTTA GTGCTGCAAC AGAAGCAGGT
GCAGCCACAG CTGCAGCAGG AGGCAGAGCC ACAGAAGCAGGTGCAGCCAC
AGGTACAGCC ACAGGCACAT TCACAGGGCC CAAGGCAGGT GCAGCTGCAG
CAGGAGGCAG AGCCGCTGAA GCAGGTGCAG CCACAGGTGC AGCCCCAGGC
ACATTCACAG CCCCCAAGGC AGGTGCAGCT GCAGCTGCAG AAGCAGGTCC
AGACACAGAC ATATCCACAG GTCCACACAC AGGCACAGCC AAGCGTCCAG
CCACAGGAGC ATCCTCCAGC GCAGGTGTCA GTACAGCCAC CAGAGCAGAC
CCATGAGCAG CCTCACACCC AGCCGCAGGT GTCGTTGCTG GCTCCAGAGC
AAACACCAGT TGTGGTTCAT GTCTGCGGGC TGGAGATGCC ACCTGATGCA
GTAGAAGCTG GTGGAGGCAT GGAAAAGACC TTGCCAGAGC CTGTGGGCAC
CCAAGTCAGC ATGGAAGAGA TTCAGAATGA GTCGGCCTGT GGCCTAGATG
TGGGAGAATG TGAAAACAGA GCGAGAGAGA TGCCAGGGGTATGGGGCGCC
GGGGGCTCCC TGAAGGTCAC CATTCTGCAG AGCAGTGACA GCCGGGCCTT
TAGCACTGTA CCCCTGACAC CTGTCCCCCG CCCCAGTGAC TCCGTCTCCT
CCACCCCTGC GGCTACCAGC ACTCCCTCTA AGCAGGCCCT CCAGTTCTTC
TGCTACATCT GCAAGGCCAG CTGCTCCAGC CAGCAGGAGT TCCAGGACCA
CATGTCGGAG CCTCAGCACC AGCAGCGGCT AGGGGAGATC CAGCACATGA
GCCAAGCCTG CCTCCTGTCC CTGCTGCCCG TGCCCCGGGA CGTCCTGGAG
ACAGAGGATG AGGAGCCTCC ACCAAGGCGC TGGTGCAACA CCTGCCAGCT
CTACTACATG GGGGACCTGA TCCAACACCG CAGGACACAG GACCACAAGA
TTGCCAAACA ATCCTTGCGA CCCTTCTGCA CCGTTTGCAA CCGCTACTTC
AAAACCCCTC GCAAGTTTGT GGAGCACGTG AAGTCCCAGG GGCATAAGGA
CAAAGCCAAG GAGCTGAAGT CGCTTGAGAA AGAAATTGCT GGCCAAGATG
```

AGGACCACTT CATTACAGTG GACGCTGTGG GTTGCTTCGA GGGTGATGAA
GAAGAGGAAG AGGATGATGA GGATGAAGAAGAGATCGAGGTTGAGGAGGA
ACTCTGCAAG CAGGTGAGGT CCAGAGATAT ATCCAGAGAG GAGTGGAAGG
GCTCGGAGAC CTACAGCCCC AATACTGCAT ATGGTGTGGA CTTCCTGGTG
CCCGTGATGG GCTATATCTG CCGCATCTGC CACAAGTTCT ATCACAGCAA
CTCAGGGGCA CAGCTCTCCC ACTGCAAGTC CCTGGGCCAC TTTGAGAACC
TGCAGAAATA.CAAGGCGGCC AAGAACCCCA GCCCCACCAC CCGACCTGTG
AGCCGCCGGT GCGCAATCAA CGCCCGGAAC GCTTTGACAG CCCTGTTCAC
CTCCAGCGGC CGCCCACCCT CCCAGCCCAA CACCCAGGAC AAAACACCCA
GCAAGGTGAC GGCTCGACCC TCCCAGCCCC CACTACCTCG GCGCTCAACC
CGCCTCAAAA CCTGATAGAG GGACCTCCCT GTCCCTGGCC TGCCTGGGTC
CAGATCTGCT AATGCTTTTT AGGAGTCTGC CTGGAAACTT TGACATGGTT
CATGTTTTTA CTCAAAATCC AATAAAACAA GGTAGTTTGG CTGTGCAAAA
AAAAAAAAAA AAAAAAAAAA AA

**Figure 8c; Human full length protein**

MF SQQQQQLQQQ QQQLQQLQQQ QLQQQQLQQQ QLLQLQQLLQQSPPQ
APLPM AVSRGLPPQQ PQQPLLNLQG TNSASLLNGS MLQRALLLQQLQ GL
DQFAMP PATYDTAGLT MPTATLGNLR GYGMASPGLA APSLTPPQLATPN
LQQFFPQ ATRQSLLGPP PVGVPMNPSQ FNLSGRNPQK QARTSSSTTPNRK
DSSSQTM PVEDKSDPPE GSEEAAEPRM DTPEDQDLPP CPEDIAKEKRTPA
PEPEPCE ASELPAKRLR SSEEPTEKEP PGQLQVKAQP QARMTVPKQTQTP
DLLPEAL EAQVLPRFQP RVLQVQAQVQ SQTQPRIPST DTQVQPKLQK
QAQTQTSPEH LVLQQKQVQP QLQQEAEPQK QVQPQVQPQAHSQGPRQ
VQLQQEAEPLKQV QPQVQPQAHS QPPRQVQLQL QKQVQTQTYP QVHT
QAQPSVQPQEHPPAQV SVQPPEQTHE QPHTQPQVSL LAPEQTPVVV HVC
sGLEMPPDAVEAGGGMEK TLPEPVGTQV SMEEIQNESA CGLDVGECEN
RAREMPGVWGAGGSLKVTIL QSSDSRAFST VPLTPVPRPS DSVSSTPAAT
STPSKQALQFFCYICKASCS SQQEFQDHMS EPQHQQRLGE IQHMSQACLL
SLLPVPRDVLETEDEEPPPR RWCNTCQLYY MGDLIQHRRT QDHKIAKQSL
RPFCTVCNRYFKTPRKFVEH VKSQGHKDKA KELKSLEKEI AGQDEDHFIT
VDAVGCFEGDEEEEEDDEDE EEIEVEEELC KQVRSRDISR EEWKGSETYS
PNTAYGVDFLVPVMGYICRI CHKFYHSNSG AQLSHCKSLG HFENLQKYKA
AKNPSPTTRPVSRRCAINAR NALTALFTSS GRPPSQPNTQ DKTPSKVTAR
PSQPPLPRRSTRLKT